The Report Committee for Rachael Celia Gilbert
Certifies that this is the approved version of the following report:

Recognition memory in noise for speech of varying intelligibility

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Rajka Smiljanic

Scott Myers

Recognition memory in noise for speech of varying intelligibility

by

Rachael Celia Gilbert, B.A.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May 2013

Acknowledgements

I would like to thank Rajka Smiljanic, Bharath Chandarasekaran, and Kristin Van Engen for their assistance with experimental design and analysis. I would also like to thank Lauren Franklin and Johanna Boardman for their assistance with data collection.

Abstract

Recognition memory in noise for speech of varying intelligibility

Rachael Celia Gilbert, M.A.

The University of Texas at Austin, 2013

Supervisor: Rajka Smiljanic

This study investigated the extent to which noise impacts speech processing of sentences that vary in intelligibility for normal-hearing young adults. Intelligibility and recognition memory in noise were examined for conversational and clear speech sentences recorded in quiet (QS) and in response to the environmental noise, i.e. noise adapted speech (NAS). Results showed that 1) increased intelligibility through conversational-to-clear speech modifications lead to improved recognition memory and 2) NAS presented a more naturalistic speech adaptation to noise compared to QS, leading to more accurate word recognition and better sentence recall. These results demonstrate that acoustic-phonetic modifications implemented in listener-oriented speech enhance speech processing beyond word recognition. The results are in line with the effortfulness hypothesis (McCoy et al., 2005), which states that speech perception in challenging listening environments requires additional processing resources that might otherwise be available for encoding speech in memory. This resource reallocation may be offset by

speaking style adaptations on the part of the talker. In addition to enhanced intelligibility, a substantial improvement in recognition memory can be achieved through speaker adaptations to the environment and to the listener when in adverse conditions.

Table of Contents

Introduction

Most communicative environments involve some degree of background noise. In everyday listening situations, accurate speech perception relies on the capacity of the auditory system to process degraded speech signals. Successful perception of speech in such adverse listening conditions requires stable sensory representations and considerable cognitive effort to extract the signal from noise. This task is challenging even for listener groups with normal hearing and normal cognitive abilities (Assmann & Summerfield, 2004; Rogers et al., 2006). The current paper examines the impact of noise on word recognition and sentence recognition memory for speech that varies in intelligibility. Specifically, we test whether speaking style adaptations aimed at improving communicative effectiveness improve recognition memory for spoken sentences that are processed under challenging listening conditions.

Extensive research has shown that noise can have a range of adverse effects on speech processing (Mattys et al., 2009; Summers et al., 1988; Junqua 1996, Assman and Summerfield, 2004). In terms of memory, Murphy et al (2000) found that digitized word pairs were significantly more difficult to recall in 12-talker babble than in quiet at both low (-5 db SNR) and moderate (-10 db SNR) noise levels. Using sentences masked by 8-talker babble, Pichora-Fuller et al. (1995) found that an increase in noise level from +5 db SNR to 0 db SNR resulted in a significant drop in word recall for both younger and older adults. In another study, young adults exhibited poorer recall for spoken digits masked by narrowband noise compared to digits in quiet, even when these same digits were accurately identified (Rabbitt, 1968). Similarly, hearing-impaired older adults were found to recall fewer words compared to listeners with typical hearing, even though they identified words equally well (Rabbitt, 1991). These studies suggest that listening to speech that is difficult to process (due to hearing impairment or noise) adversely affects memory by detracting from encoding and rehearsal abilities. These provide support for the effortfulness hypothesis: the idea that additional effort needed for perceptual processing in challenging listening environments costs processing resources that might otherwise be dedicated to the encoding of information in memory (mccoy et al., 2005). The corollary of this notion is that releases in the perceptual efforts required to successfully recognize a degraded signal should free up more processing resources for memory-related tasks.

It is well established that listener-oriented styles of speech production can enhance intelligibility, providing a release in perceptual effort (Smiljanic and Bradlow, 2005). Talkers naturally adopt a “clear” speaking style when they are aware of a speech perception difficulty on the part of the listener (e.g. Hearing impairment or low proficiency in the language). In line with the definition used in previous research, we use the term clear speech to refer to read laboratory speech elicited by instructions given to talkers rather than to the spontaneous speech occurring in a more natural setting (for a discussion of terminology, see Smiljanic and Bradlow, 2009). The instructions most typically involve asking talkers to read the same set of materials twice: once in conversational and once in clear speaking style (Picheny et al. 1986; Schum 1996; Krause and Braida 2002; Ferguson and Kewley-Port 2002; Ferguson 2004; Smiljanic and Bradlow 2005; Smiljanic and Bradlow, 2009). The resulting clear speech involves a decrease in speaking rate (longer segments as well as longer and more frequent pauses), a wider dynamic pitch range, greater sound-pressure levels, more salient stop releases, an expanded vowel space, greater obstruent RMS energy, and increased energy in 1000-3000 Hz range of long-term spectra (Smiljanic and Bradlow, 2005; Picheny et al., 1986; Krause and Braida, 2004; Bradlow et al., 2003; Liu et al., 2004, Ferguson and Kewley-

Port, 2002). The clear speech benefit has been documented for listeners in different languages, for young and old adults, for native and nonnative listeners, and for listeners with hearing impairments (Picheny et al., 1985; Smiljanic and Bradlow, 2005; Bradlow and Bent, 2002; Bradlow et al., 2003; Liu et al., 2004).

The majority of these studies investigating the perceptual processing of speech in noise have used speech signals that were recorded under quiet conditions. This may be an issue because, in the words of Chung et al. (2005), "the speech signal presented to perceivers has no correlation with the noise accompanying it and may, therefore, introduce yet one more distortion in the already strained realism of laboratory test conditions." Speech produced in the presence of actual noise, or noise-adapted speech (NAS, i.e. Lombard speech), has a number of acoustic/articulatory characteristics that differentiate it from speech produced in quiet conditions. Lombard speech is characterized by an increase in loudness, vowel duration, fundamental frequency, as well as a flattening of spectral tilt (Lombard, 1911; Summers et al., 1988; Junqua, 1993; Lane and Tranel, 1971). Previous research has shown that noise-adapted speech is more intelligible in noise than speech recorded in quiet and mixed with noise, both for native and nonnative listeners (Dreher and O'Neill, 1957; Pittman and Wiley, 2001; Summers et al. 1988; Cooke and Lecumberri, 2012).

While the beneficial effects of these two intelligibility-enhancing speaking style modifications are well established in terms of word recognition, very little is known about how variation in intelligibility impacts the encoding of speech in memory. In a recent study, Van Engen et al., (2012) found that young adults showed better recognition memory for clear speech sentences relative to conversational sentences in quiet. However, the extent to which this benefits applies to processing speech in noise has not been established thus far.

To this end, this study investigated the extent to which intra-talker variation in speech clarity impacted recognition memory for sentences in noise. We also compared the processing of speech recorded in quiet and mixed with noise with that of speech recorded in response to the environmental noise. In experiment 1 we examined the intelligibility of conversational and clear speech produced in quiet and in the presence of noise for normal-hearing, native adult speakers of English. In experiment 2, we tested the extent to which listener-oriented conversational and clear speech produced in quiet and in response to noise affect recognition memory for sentences presented in noise. Acoustic analyses were also performed on the sentences in order to examine the acoustic-phonetic changes characteristic of clear speech and noise-adapted speech. The results provide one of the first direct comparisons between listener- and environment-oriented acoustic-phonetic enhancements, as well as their effects on speech intelligibility. Furthermore, this is one of the first studies to investigate the extent to which clear speech and noise-adapted speech benefits extend to speech processing tasks beyond word recognition.

Per the effortfulness hypothesis, we predicted that for sentences in which perceptual effort is reduced (through acoustic-phonetic enhancements associated with clear speech and noise-adapted speech), more resources would be available for encoding speech in memory. Thus we expected both clear and noise-adapted speech modifications to enhance recognition memory for listeners. We also predicted that clear speech produced in response to noise would show a cumulative processing benefit, i.e. That noise-adapted clear speech would be the most intelligible and provide the largest recognition memory boost. Our results revealed that these intra-talker speaking style adaptations do indeed improve speech-in-noise intelligibility (Experiment 1) and enhance recognition memory in noise (Experiment 2). These results reveal that improving speech clarity reduces processing demands imposed by adverse listening conditions and allows for better encoding of speech in memory even in noise. In sum, the resource reallocation induced by listening in adverse situations may be offset by speaker adaptations to the environment and to the listener.

Experiment I: Intelligibility

MATERIALS

The stimuli consisted of 80 meaningful sentences modified from the Basic English Lexicon (BEL) sentence materials (Calandruccio and Smiljanic, 2012) and used in Van Engen et al. (2012). Sentences each contained four keywords for intelligibility scoring (e.g. *The <u>small</u> <u>animal</u> <u>scared</u> the <u>baby</u>)*. One native female speaker of American English (aged 26 years, with no speech or hearing impairment) was recorded producing the full set of 80 meaningful sentences over two sessions.

The two sessions differed in the type of talker response to the environment. In the first session, the speaker read the stimuli *in quiet* (quiet speech; QS). In the second session, the speaker read the same sentences *in the presence of 6-talker babble* presented via Sennheiser HD5 headphones (noise-adapted speech; NAS). The 6-talker babble was composed of six monolingual speakers of American English (three females and three males between the ages of 28 and 48 years) producing semantically anomalous sentences in English (Van Engen and Bradlow, 2007; Smiljanic and Bradlow, 2005). Anomalous sentences were used to minimize chance that listeners might extract a meaningful sentence other than the target. The NAS and QS sessions took place six months apart. Both recording sessions took place in a sound-attenuated booth.

In each session, all sentences were read once in a conversational speaking style and once in a clear speaking style. Conversational speech was elicited by instructing the speaker to speak in a casual, conversational style, as if she was talking to someone familiar with her voice and speech patterns. The clear speech was elicited by instructing her to speak as though the listener was having a hard time understanding her, whether due to hearing difficulty or because the listener was a non-native speaker of English. For the speech produced in noise, the same instructions for eliciting the two listener-oriented speaking styles were given as in quiet. These instructions have been shown to be sufficient for elicitation of different speaking styles (for a review of clear speech and Lombard speech elicitation instructions, see Pichora-Fuller et al., 2010 or Smiljanic and Bradlow, 2009).

Sentences were presented to the speaker one at a time on a computer monitor. Recordings were made using a Shure SM10A head-mounted microphone and a MOTU UltraLite-MK3 Hybrid recorder. The recorded sentences were segmented into individual files which were equalized for RMS amplitude across the entire sentence duration. The total set of recorded sentences was 320 (80 conversational QS; 80 clear QS; 80 conversational NAS; 80 clear NAS).

LISTENERS

Sixteen adults between the ages of 18 and 34 served as listeners. All participants were native, monolingual speakers of American English. All were University of Texas at Austin undergraduate students. They all passed a hearing-screening test (1000, 2000, and 4000 Hz at 25 dB). Participants provided written informed consent and were either paid for their participation or received course credit.

PROCEDURE

All sentences were presented to listeners for assessment of intelligibility. Speech-shaped noise (SSN) was created for each sentence type (conversational QS, clear QS, conversational NAS, and clear NAS) by filtering white noise to the long-term average spectrum of the full set of sentences. This approach was used to ensure that masking was constant across the recording types (following Van Engen et al., 2012). Each file was digitally mixed with noise at a signal-to-noise ratio (SNR) of -5 dB SPL. The SNR of -5 dB was determined from pilot testing to ensure that listeners would not perform at the ceiling in the easiest listening condition. Each of the final stimulus files consisted of a 400 ms silent lead, followed by 500 ms of noise, followed by the speech-plus-noise files, and ending with 500 ms of only noise. The noise preceding and following the speech stimulus was at the same level as the noise mixed with the speech.

The session began with a language background questionnaire that detailed the participants' language learning experiences and general education. Participants were then brought into a sound-attenuated booth and screened for normal hearing. For the intelligibility test, each participant was seated in front of a computer monitor. The stimuli were played over headphones (Sennheiser HD570 or Sony MDR-CD780) at a comfortable listening level set by the experimenter. Instructions and stimuli were presented using EPrime (Schneider et al., 2002). The participant's task was to listen to each sentence and write down as much as they could onto a prepared answer sheet. After each trial, the participant pressed a button on the keyboard to move onto the next trial. Each trial was presented only once, but participants could take as much time as they wished to write down the sentences.

In order to familiarize participants with the task, the session began with four practice items not included in the subsequent test. Each participant then transcribed a total of eighty pseudorandomized sentences from either the QS recordings or the NAS recordings. Forty of these sentences were produced in conversational speech, and forty were produced in clear speech; this was counterbalanced so that half of the participants heard sentences produced in the opposite speaking styles as the other half of the participants. Participants never heard the same sentence twice. Each sentence was scored by the number of keywords correctly identified (4 per sentence) for a total of 160 keywords per sentence type per listener. In order to be considered correct, no morphemes could be added to or deleted from the keywords, but homophones were acceptable.

RESULTS

Intelligibility results showed that, for QS, listeners identified 38.71% of the keywords in conversational speech (SD: 9.81%) and 77.19% of the keywords in clear speech (SD: 7.43%). For NAS, listeners correctly identified 52.25% of the keywords in conversational speech (SD: 9.03%) and 86.81% of the keywords in clear speech (SD: 5.46%). Proportion correct scores for conversational and clear sentences produced in QS and NAS are shown in Figure 1.

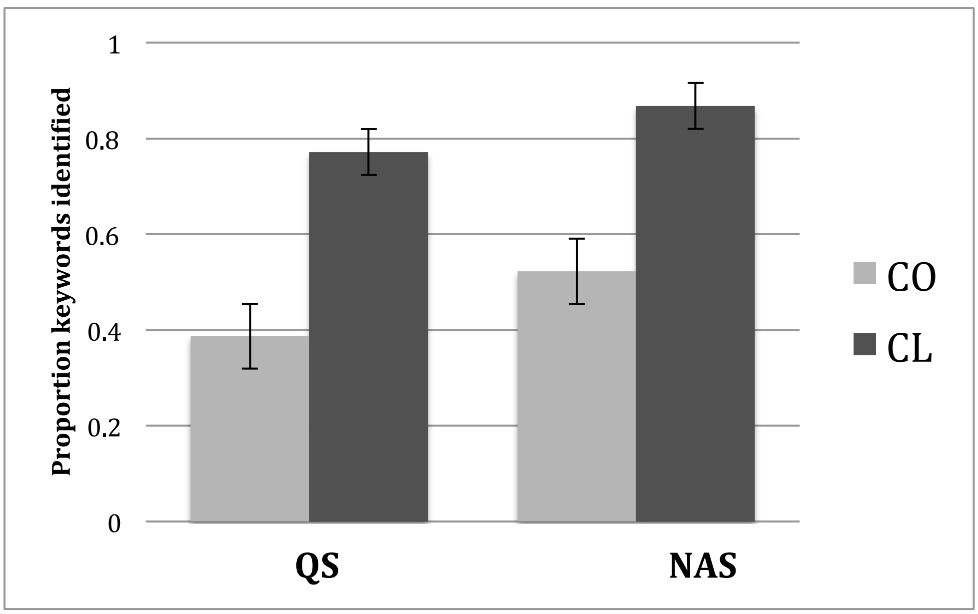

Figure 1: Average proportion of keywords identified from conversational and clear sentences produced in quiet (QS) and in noise (NAS). Error bars represent standard error.

The intelligibility data were analyzed with a linear mixed effects logistic regression where keyword identification (i.e. correct or incorrect) was the dichotomous dependent variable. Listener and Sentence were included in the model as random factors and Listener-Oriented Speaking Style (conversational or clear), Environment-Oriented Speaking Style (QS or NAS), and their interactions as fixed effects. Listener-Oriented Speaking Style was contrast coded (-0.5, 0.5) such that negative beta values are associated with clear speech and positive beta values are associated with conversational speech. Environment-Oriented Speaking Style was also contrast coded (-0.5, 0.5) such that negative beta values are associated with speech produced in response to noise and positive beta values are associated with speech produced in quiet. The results of the regression are presented in Table 1.

Fixed effects:	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	0.948	0.126	7.520	5.28e-14***
Environment-Oriented Speaking Style	-2.120	0.056	-38.180	<2e-16***
Listener-Oriented Speaking Style	-0.998	0.155	-6.460	1.08e-10***
Enviroment-Oriented SS:Listener Oriented SS	0.237	0.107	2.230	0.0259*

Table 1: Results of the linear mixed effects logistic regression on intelligibility data for all sentences.

The results revealed that the overall probability of correct keyword identification was significantly higher for NAS versus QS speech ($p<0.001$) and for clear versus conversational speech ($p<0.001$). Results also revealed a significant interaction between Listener-Oriented Speaking Style and Environment-Oriented Speaking Style ($p=0.026$). The nature of this interaction was examined by performing a second round of mixed-effects logistic regressions on the QS and NAS sets individually. The results of these regressions are shown in Tables 2 and 3. Although the effect of listener-oriented speaking style was a highly significant predictor of correct keyword identification for both QS and NAS, the effect of style was greater for NAS ($\beta_{QS} = -2.033$ whereas $\beta_{NAS} = -2.291$).

Fixed effects:	**Estimate**	**Std. Error**	**z value**	**Pr(>lzl)**
(Intercept)	0.454	0.131	3.455	0.000551***
Listener-Oriented Speaking Style	-2.033	0.071	-28.790	<2e16***

Table 2: Results of the linear mixed effects logistic regression on intelligibility data for QS sentences.

Fixed effects:	**Estimate**	**Std. Error**	**z value**	**Pr(>lzl)**
(Intercept)	1.503	0.178	8.421	<2e-16***
Listener-Oriented Speaking Style	-2.291	0.088	-25.948	<2e-16***

Table 3: Results of the linear mixed effects logistic regression on intelligibility data for NAS sentences.

These findings support previous evidence showing that listener-oriented conversational-to-clear speech modifications enhance sentence intelligibility (Smiljanic and Bradlow, 2009). Furthermore, these results not only confirm that QS artificially mixed with noise is more challenging for listeners than naturally produced NAS (Dreher and O'Neill, 1957; Pittman and Wiley, 2001; Summers et al. 1988), but also reveal that the clear speech gain under adverse listening conditions, i.e. in noise, is larger for NAS than for QS. This reflects the difference in the talker response to the environment, i.e. producing deliberate clear speech in quiet vs. in response to the actual background babble. Experiment 2 investigates the extent to which these differences in intelligibility impact recognition memory.

Experiment II: Recognition memory

MATERIALS

Forty sentences each in conversational and clear speech from the QS and NAS conditions (160 total) were used in the recognition memory experiment. This subset of the total sentences from Exp 1 was selected due to experimental duration and memory load constraints. The QS sentences were also the same meaningful sentences used in Van Engen's recognition memory experiment (Van Engen et al., 2012). Results from Van Engen et al. showed that these QS sentences used as old and new for recognition memory did not vary systematically in their intelligibility. In order to confirm this pattern for NAS sentences, an unpaired, 2-tailed t-test was conducted on the NAS intelligibility data for these sentences. The test showed no significant difference between the intelligibility of NAS old and new sentences either ($p=0.0511$).

LISTENERS

Sixty native speakers of American English between the ages of 18 and 30 participated in the experiment. They were drawn from the same participant pool as in Experiment 1. The same inclusion criteria applied as in Experiment 1. Participants provided written informed consent and were either paid for their participation or received course credit.

PROCEDURE

The same 6-talker babble file played through headphones to the speaker during the production of NAS served as the noise masker in the exposure phase. To avoid masker familiarization, each meaningful sentence was digitally mixed with one of four different portions of the babble file. Sentences were mixed at signal-to-noise ratios of 0 and +3 dB SPL. The SNRs of 0 and +3 dB were determined from results of Van Engen et al (2012) and pilot testing. Each of the final stimulus files consisted of a 400 ms silent lead, followed by 500 ms of noise, followed by the speech-plus-noise files, and ending with 500 ms of only noise. The noise preceding and following the speech stimulus was at the same level as the noise mixed with the speech.

Testing setup was the same as in Experiment 1. Participants took part in one of four experimental conditions: recognition memory for conversational and clear QS mixed with noise at 0 dB SNR (n=15), for conversational and clear QS mixed with noise at +3 dB SNR (n=15), for conversational and clear NAS mixed with noise at 0 dB SNR (n=15), and for conversational and clear NAS mixed with noise at +3 dB SNR (n=15). In each experimental condition, listeners were first exposed to 40 unique sentences embedded in the 6-talker babble and instructed to try to commit them to memory (the exposure phase). Twenty of the sentences were presented in conversational speech and 20 in clear speech. Sentences were presented only once with a 500 ms break between sentences. In the test phase, participants were instructed to listen to a second set of sentences and indicate by pressing one of two buttons whether each sentence was old (from the exposure phase) or new. All 40 of the exposure sentences were included along with 40 new sentences. Half the sentences were in conversational speech and half were in clear speech. Sentences in the test phase were presented in quiet. The participants were instructed to press a third button on the response box to move from trial to trial. Each trial was presented only once. In both phases, sentence order was randomized for each participant.

RESULTS

Recognition accuracy was analyzed using the *d'* statistic in order to provide a measure of accuracy independent of response bias (Lamont 2005). The *d'* statistic accounts for discrimination sensitivity and bias in each participant by subtracting the normalized probability of false alarms (i.e. identifying a new sentence as old) from the normalized probability of hits (i.e. identifying an old sentence as old), and then correcting the formula to account for values of 0 and 1. Table 4 lists all normalized hit rates, false alarm rates, d' statistics, and C scores (a measure of changes in response bias). The average C scores across all conditions are positive, meaning that participants were generally biased to respond "new" more often than "old". This bias was stronger for clear speech than for conversational speech. The overall results of Experiment 2 are shown in Figure 2.

		Conversational Speech				**Clear speech**			
	SNR	**Hit Rate**	**False Alarm Rate**	**d'**	**C**	**Hit Rate**	**False Alarm Rate**	**d'**	**C**
QS	**0**	0.51	0.31	0.55	0.24	0.58	0.20	1.14	0.36
	3	0.60	0.31	0.76	0.13	0.64	0.26	1.12	0.19
NAS	**0**	0.60	0.20	1.17	0.34	0.66	0.14	1.56	0.33
	3	0.64	0.26	1.12	0.18	0.58	0.17	1.26	0.40

Table 4: Normalized hit rates, false alarm rates, d', and C values for the recognition memory test for sentences produced in quiet (QS) and in noise (NAS).

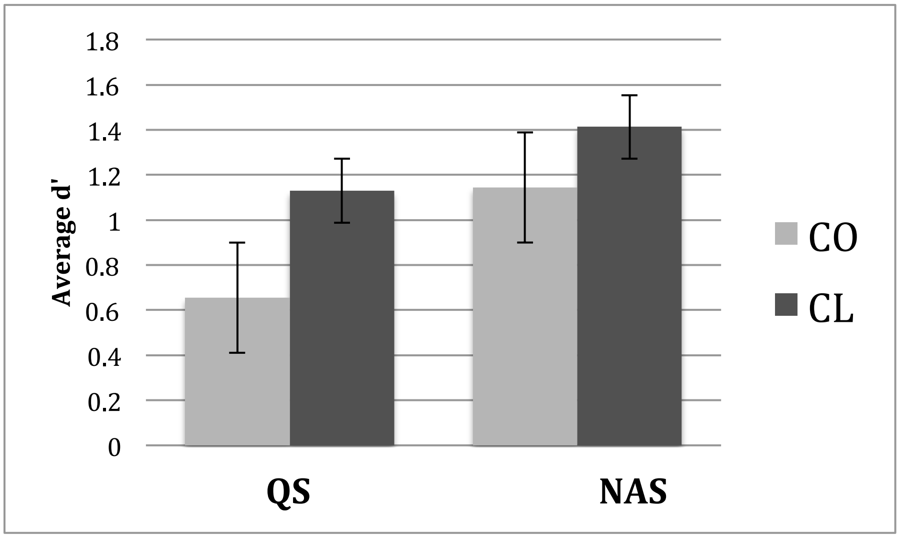

Figure 2: Average d' scores for conversational and clear sentences produced in quiet (QS) and in noise (NAS). Error bars represent standard error.

D' scores were submitted to a mixed ANOVA with Listener-Oriented Speaking Style (conversational or clear) as the within-subject factor and Environment-Oriented Speaking Style (QS or NAS) and SNR (0 dB or +3 dB) as between-subject factors. There was a main effect of Listener-Oriented Speaking Style ($F(1,56)=22.310$, $p<0.001$) and of Environment-Oriented Speaking Style ($F(1,1)=10.223$, $p=0.002$) on d' scores. The effect of SNR (0 dB vs. +3 dB) was not significant ($F(1,1)=0.105$, $p=0.747$). No significant interactions between Listener-Oriented Speaking Style, Environment-Oriented Speaking Style, and SNR were found (Listener-Oriented Speaking Style by Environment-Oriented Speaking Style: $F(1,1)=1.718$, $p=0.195$, Listener-Oriented Speaking Style by SNR: $F(1,1)=2.261$, $p=0.138$, Environment-Oriented Speaking Style by SNR: $F(1,1)=1.277$, $p=0.263$, Listener-Oriented Speaking Style by Environment-Oriented Speaking Style by SNR: $F(1,1)=0.010$, $p=0.920$).

The results showed that speech clarity significantly contributed to listeners' enhanced recognition memory for sentences. Listeners were better able to recognize previously heard sentences when they were produced in clear speech relative to conversational speech and in NAS relative to QS. This effect was significant even though listeners were committing these sentences to memory in noise. Finally, the results showed that recognition memory scores did not significantly differ across QS and NAS clear speech, unlike in experiment 1.

Acoustic Analyses

Procedure

Four specific acoustic-articulatory features were measured for each recorded sentence: speech rate (syllables per second), F0 range (Hz), mean F0 (Hz), and energy in the 1-3 kHz range (dB). All acoustic features were measured per sentence and then averaged to obtain listener-oriented (conversational, clear) and environment-oriented (QS, NAS) speaking style values. Speech rate was calculated as the number of syllables divided by the sentence duration, excluding pauses greater than 100 ms. F0 range was calculated as the difference between the sentence's minimum F0 value and peak F0 value. Mean F0 was an average of F0 values over the entire sentence. Energy in the 1-3 kHz range was measured by averaging the long-term average spectrum energy between 1 and 3 kHz across the sentence. We focused on these temporal-, pitch-, and spectral-related features, as they are typical metrics for contrasting conversational speech against clear speech as well as quiet speech against Lombard speech (Pichora-Fuller et al., 2010; Smiljanic and Bradlow, 2009).

Results

A series of acoustic analyses was performed on all sentences in order to assess whether the two listener-oriented speaking styles (conversational vs. clear) and the two environment-oriented speaking styles (QS vs. NAS) differed in their acoustic-articulatory characteristics. The results of four acoustic measurements (speaking rate (syllable/second), F0 mean and range (Hz), and energy in the 1-3 kHz range (dB)) are given in Table 5. Waveforms and spectrograms of an example sentence in each listener- and environment-oriented speaking style are shown in in ure 3. For each of the four measurements, results were submitted to a repeated measures ANOVA with Listener-

Oriented Speaking Style (conversational or clear) and Environment-Oriented Speaking Style (QS or NAS) as the within-sentence factors.

Mean (SD)	QS CO		QS CL		NAS CO		NAS CL	
Speech Rate (syllables/sec)	5.30 (0.98)	5.14 (0.92)	2.81 (0.52)	2.79 (0.53)	4.75 (0.64)	4.66 (0.61)	2.27 (0.50)	2.29 (0.57)
Average F0 (Hz)	161.24 (8.85)	161.19 (9.22)	167.42 (7.90)	166.33 (7.95)	179.49 (5.78)	179.66 (5.21)	198.22 (7.32)	198.04 (7.38)
F0 range (Hz)	136.79 (108.25)	137.96 (108.71)	215.63 (122.69)	210.69 (119.18)	117.94 (38.44)	123.05 (43.46)	179.63 (60.52)	179.59 (59.01)
Energy: 1-3 kHz	22.61 (2.52)	22.64 (2.35)	22.17 (2.63)	22.83 (2.73)	28.60 (1.76)	28.76 (1.80)	28.66 (1.72)	28.98 (1.77)

Table 5: Acoustic measures of sentence materials as produced in quiet (QS), in noise (NAS), in conversational speech (CO), and in clear speech (CL). Within each cell, the number on the left represents the average for all sentences (Experiment I) and the number on the right represents the average for the subset of sentences used in Experiment II.

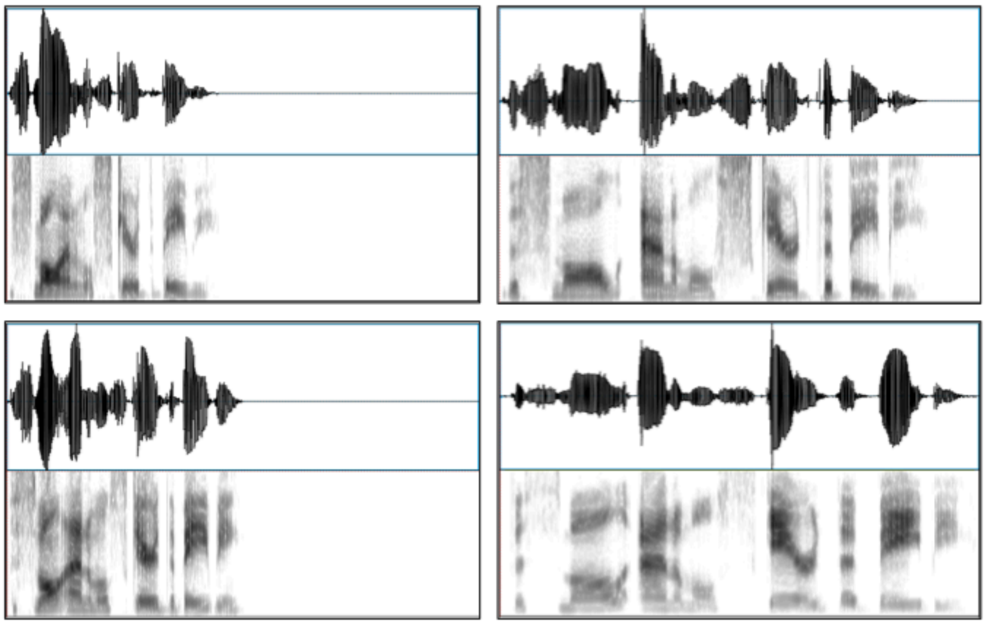

Figure 3: Waveforms and spectrograms of one sentence (*The small animal scared the baby*) produced in quiet (top panels) and in noise (bottom panels), each produced in both conversational (left panels) and clear (right panels) speaking styles. Each panel display represents 3.42 seconds. Notice the differences in duration between speech produced in quiet and in noise (top vs. bottom), and, to an even greater extent, conversational and clear speech (left vs. right).

Speaking rate

For speaking rate, there were main effects of Listener-Oriented Speaking Style ($F(1,159)=3050.756$, $p<0.001$) and Environment-Oriented Speaking Style ($F(1,159)=279.977$, $p<0.001$), with clearly produced sentences showing significantly slower speech rates than sentences in conversational speech, and sentences produced in noise showing significantly slower speech rates than sentences produced in quiet. There was no significant interaction between Listener-Oriented Speaking Style and Environment-Oriented Speaking Style ($F(1,159)=0.025$, $p=0.875$).

In order to ensure that these acoustic trends held for sentences used in both experiments, this set of analyses was repeated on just the subset of sentences used in Experiment II. Experiment II sentences were characterized by the same speaking rate patterns, showing main effects of Listener-Oriented Speaking Style ($F(1,79)=1366.233$, $p<0.001$) and Environment-Oriented Speaking Style ($F(1,79)=93.143$, $p<0.001$), with no significant interaction between Listener-Oriented Speaking Style and Environment-Oriented Speaking Style ($F(1,79)=0.032$, $p=0.858$).

F0 Range

Significant main effects of Listener-Oriented Speaking Style and of Environment-Oriented Speaking Style were found for F0 range: ($F(1,159)=100.200$, $p<0.001$), ($F(1,159)=14.233$, $p<0.001$). Sentences in clear speech showed significantly larger F0 ranges than sentences in conversational speech, whereas noise-adapted sentences showed significantly smaller F0 ranges than their counterpart sentences produced in quiet. There was no significant interaction ($F(1,159)=1.719$, $p=0.192$).

This set of analyses was also repeated on just the subset of sentences used in Experiment II. The second round of analyses confirmed that the sentences in both experiments shared the same F0 range characteristics, with the subset showing significant main effects of Listener-Oriented Speaking Style and of Environment-Oriented Speaking Style: ($F(1,79)=44.043$, $p<0.001$), ($F(1,79)=5.682$, $p=0.020$). Again, there was no significant interaction ($F(1,79)=0.668$, $p=0.416$).

F0 Mean

There were significant main effects of Listener-Oriented Speaking Style and Environment-Oriented Speaking Style on average F0: ($F(1,159)=527.344$, $p<0.001$), ($F(1,159)=2339.655$, $p<0.001$). Sentences in clear speech and noise-adapted speech exhibited significantly higher mean F0 than sentences in conversational speech and speech produced in quiet, respectively. A significant interaction between Listener-Oriented Speaking Style and Environment-Oriented Speaking Style was also found ($F(1,159)=140.765$, $p<0.001$), with noise-adapted clear speech exhibiting a significantly higher mean F0 than could be attributed to the cumulative effects of Lombard speech and clear speech alone.

Analyses on just the subset of sentences used in Experiment II showed the same acoustic trends, with significant main effects of Listener-Oriented Speaking Style ($F(1,79)=215.032$, $p<0.001$) and Environment-Oriented Speaking Style ($F(1,79)=1237.327$, $p<0.001$), as well as a significant interaction between Listener-Oriented Speaking Style and Environment-Oriented Speaking Style ($F(1,79)=72.987$, $p<0.001$).

Energy in the 1-3 kHz region

There was a significant main effect of Environment-Oriented Speaking Style on energy in 1-3 kHz range, ($F(1,159)=2736.700$, $p<0.001$); sentences adapted to noise were characterized by significantly greater energy in this range than sentences produced in quiet. This trend was present but marginally significant for Listener-Oriented Speaking Style ($F(1,159)=3.676$, $p=0.057$). A significant interaction between Listener-Oriented Speaking Style and Environment-Oriented Speaking Style was found as well ($F(1,159)=10.819$, $p=0.001$); for sentences produced in quiet, conversational speech exhibited slightly more 1-3 KHz energy than clear speech, whereas the inverse held true for sentences adapted to noise. This is unusual, given that clear speech in quiet typically exhibits higher energy in the 1-3 kHz range than does conversational speech (Van Engen et al, 2012).

For the subset of sentences used in Experiment II, there was again a significant main effect of Environment-Oriented Speaking Style on energy in 1-3 kHz range, ($F(1,79)=1353.628$, $p<0.001$). Again, the effect of Listener-Oriented Speaking Style did not reach significance ($F(1,79)=2.237$, $p=0.139$). No significant interaction was found ($F(1,79)=0.022$, $p=0.883$).

The analyses of sentences used in Experiment I showed that sentences produced in a clear speaking style were overall slower, had higher F0 means, and exhibited wider F0 ranges compared to sentences produced in conversational speech. Speaking in response to the environmental noise (NAS) lead to slower speaking rates, higher F0 means, and greater energy in the 1-3 kHz range compared to speech produced in quiet (QS). Unexpectedly, NAS sentences showed significantly smaller F0 ranges compared to sentences produced in quiet (c.f. Jessen et al., 2005). Although non-significant, conversational QS exhibited slightly more 1-3 KHz energy than clear QS, whereas the inverse pattern held true for conversational and clear NAS (following the trend observed in Van Engen et al, 2012). Finally, the cumulative effect of listener- and environment-oriented speaking styles was manifested in clear NAS exhibiting a significantly higher mean F0 than clear QS.

The subset of sentences used in Experiment II exhibited very similar acoustic trends: sentences produced in a clear speaking style were again significantly slower, had higher F0 means, and exhibited wider F0 ranges compared to sentences produced in conversational speech. Speaking in response to the environmental noise (NAS) again lead to slower speaking rates, higher F0 means, and greater energy in the 1-3 kHz range compared to speech produced in quiet (QS). As in the overall set of stimuli, the Experiment II NAS sentences showed significantly smaller F0 ranges compared to their counterpart sentences produced in quiet. A significant interaction was again present with clear NAS exhibiting a significantly higher mean F0 than clear QS. The subset only exhibited slight acoustic-phonetic differences: the marginal effect of Listener-Oriented Speaking Style on energy in the 1-3 kHz range was even less significant in the Experiment II stimuli. Additionally, the unusual interaction between Listener- and Environment-Oriented Speaking Styles on energy in the 1-3 KHz range was not present in the subset. Ultimately, the subset of sentences used in Experiment II very closely resembled the acoustic patterns present in the overall set of sentences.

These analyses thus confirmed that the clear speech and NAS adaptations exhibited acoustic-articulatory characteristics commonly observed in listener- and environment-oriented speaking style adaptations (Pichora-Fuller et al., 2010; Smiljanic and Bradlow, 2009). Furthermore, the results demonstrated that speaking clearly and in response to environmental noise separately and in combination contribute to enhancement of acoustic-articulatory features such as pitch.

Discussion

This study examined the extent to which speaking style adaptations to the environment and to the listener facilitate word recognition and recognition memory for spoken sentences in noise. Experiment 1 tested the intelligibility of sentences produced in clear and conversational styles, recorded in quiet and in response to 6-talker babble. Experiment 2 evaluated recognition memory for these four types of speech. The acoustic properties of the sentences were also examined. The results showed that listener- and environment-oriented acoustic-phonetic enhancements to the speech signal resulted in increased intelligibility, as evidenced by improved word recognition in noise (Experiment 1). This intelligibility-in-noise gain was greater for clear speech when it was produced in response to environmental noise than when it was produced in quiet. Importantly, the current findings provide new evidence that clear speech and noise-adapted speech benefits extend to better sentence recognition memory through enhanced encoding of speech in noise compared to conversational speech and speech produced in quiet (Experiment 2).

The results of Experiment 1 are in accordance with previous research showing that clear speech enhances intelligibility in response to perceptual difficulties on the part of the listener (Uchanski, 2005; Picheny et al., 1985; Smiljanic & Bradlow, 2009) and that NAS is more intelligible than speech recorded in quiet and artificially mixed with noise (Dreher and O'Neill, 1957; Pittman and Wiley, 2001; Summers et al. 1988). To our knowledge, this is one of the first direct comparisons of the two listener- and environment-oriented speaking styles. We found that deliberately-produced clear speech in response to the environmental noise impacted intelligibility in a cumulative manner. That is, the intelligibility benefit from clear speech was significantly greater for speech produced in noise than for speech produced in quiet. This suggests that intelligibility-enhancing adaptations, e.g. clear speech, may be negatively impacted when the listening conditions do not correlate with the conditions in which the speech was produced.

It is important to note that the clear speech in this study was produced following instructions and not spontaneously in response to an actual listener. Some research has shown that this method of elicitation for clear speech can produce slightly more extreme acoustic-phonetic changes than natural speech (Hazan and Baker, 2011), whereas others have found no significant acoustic-phonetic differences between deliberately or inadvertently produced clear speech (Bond and Moore, 1994). Replicating the experiment using recordings of different speakers with and without communicative intent will be necessary to disambiguate the extent of the influence of these methodological choices. Further research is also necessary to examine the influence of the type and level of noise in noise-adapted speech recordings, given the evidence that the characteristics of intelligibility-enhancing speech vary according to the quality of the noise masker and the needs of the speaker (Hazan and Baker, 2011; Uther et al., 2007; Burnham et al., 2002).

The results of Experiment 2 demonstrated that clear speech and NAS lead to better performance in recognition memory for sentences. Speaking style adaptations that enhanced intelligibility in Experiment 1 also enhanced recognition memory in Experiment 2. Thus, the effect of speech clarity extends beyond facilitating word recognition. This finding expands upon recent work examining the effects of clear speech on recognition memory in quiet (Van Engen et al., 2012) by showing that more intelligible speech—even when presented in noise—allows for better encoding of speech in memory. In all conditions, recognition memory for both clear speech and NAS was characterized by lower rates of false alarm responses (i.e. identifying new sentences as old) than their conversational speech and QS counterparts (Table 4). This replicates the pattern of clear speech results from Van Engen et al., 2012. Similar patterns have also been found in other studies (Lamont et al., 2005; Podd, 1990; Davies, 1979), and it has been suggested that the correlation between exaggerated acoustic-phonetic cues and lower false alarm rates is indicative of enhanced memory traces for more distinctive speaking styles (Van Engen et al., 2012).

In contrast to Experiment 1, Experiment 2 did not yield an interaction between listener- and environment-oriented speaking styles. While the intelligibility gain for NAS clear speech was larger compared to that for QS clear speech, this did not translate to enhancement in recognition memory. Given that the SNRs in Experiment 2 were higher than that of Experiment 1, perhaps the subtle increase in intelligibility due to the contribution of both types of speaking styles may yield recognition memory benefits in more challenging listening conditions.

Our recognition memory findings provide further support for the effortfulness hypothesis (McCoy et al., 2005), the idea that reducing the effort associated with perceptual speech processing will free up processing resources for encoding speech in memory. In this study, the more easily recognized speaking adaptations (i.e. clear speech and NAS) were better encoded in memory. The results suggest that, because these speaking styles require less effort to process, more channel capacity can be recruited for speech encoding and rehearsal. Although listening in noise is an effortful process, listener- and environment-oriented speaking styles can mitigate the processing load.

Acoustic analysis confirmed that both clear speech and NAS exhibited typical intelligibility-enhancing features (Pichora-Fuller et al., 2010; Smiljanic and Bradlow, 2009). Compared to conversational speech, clear speech exhibited slower speech rates, higher pitch, and larger F0 ranges. Compared to QS, NAS was characterized by slower speech rates, higher pitch, and more energy in the 1-3 kHz range. It is important to note that, despite the wealth of production and perception studies in clear speech and NAS research, a direct relationship between some of the acoustic-phonetic features examined and intelligibility is still rather tenuous (Ferguson, 2004; Picheny et al. 1989; Stollman et al. 1994; Uchanski et al. 1996). Furthermore, it remains to be seen whether these are the same features that contribute to the observed improvements in recognition memory. Establishing the exact mapping between acoustic-articulatory modifications and perceptual benefits thus remains a challenging question to researchers.

Conclusions

While it is well established that clear speech improves intelligibility for various listener groups and under different listening conditions (Picheny et al., 1985; Smiljanic and Bradlow, 2005, 2011; Bradlow and Bent, 2002; Bradlow et al., 2003; Liu et al., 2004), little is known about the effects of clear speech on recognition memory. The current finding that clear speech enhances recognition memory in noise provides new evidence for the beneficial effects of clear speech on speech processing beyond word recognition. These results further suggest that, in addition to a significant cross-talker variability effect (Goldinger, 1996; Palmeri et al., 1993), within-talker variability also impacts recognition memory. Finally, the results reported here that noise-adapted speech is both more intelligible and better remembered than quiet speech mixed with noise contributes to the growing area of work illustrating the need for more naturalistic speech-in-noise perception research (i.e. incorporating noise-adapted stimuli). This is one of the first studies examining recognition memory in noise with actual noise-adapted speech.

The results of this study have several practical and clinical implications. The finding that recognition memory is influenced by variability in speech intelligibility suggests that, on the one hand, simple speaking style adaptations on the part of the talker can improve listener comprehension and recognition memory in real-world, noisy situations such as in the classroom. Conversely, recognition memory may be adversely affected by other sources of variability in speech intelligibility such as speech production impairments or foreign-accented speech—any speech which requires additional cognitive effort to process. By the same token, listener groups who must expend extra effort to perceptually process speech (older adults, cochlear implant users, people with auditory processing deficits, etc.) may be similarly disadvantaged concerning encoding resources.

The results of this study highlight that memory difficulties may in part stem from perceptual difficulties and may be offset by adopting a clearer speaking style. The current findings may provide a resource for those who regularly communicate with these listener groups, including teachers, spouses, and clinicians. Finally, our finding that speech adapted to noise was more intelligible and better remembered than speech recorded in quiet and mixed with noise suggests that the majority of speech-in-noise results utilizing the latter may be overestimating the effects of noise on speech perception. Ultimately, it is important to conduct speech perception research that simulates real-world communicative conditions as best as possible.

References

Assmann, P.F., and Summerfield, Q. (2004). The perception of speech under adverse acoustic conditions. In S. Greenberg, W.A. Ainsworth, A.N. Popper, & R.R. Fay (Eds.), *Speech processing in the auditory system* (pp. 231–308). New York: Springer Verlag.

Bradlow, A. R., and Bent, T. (2002). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. 112, 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (2003). "Speaking clearly for children with learning disabilities: Sentence perception in noise," J. Speech Lang. Hear. Res. 46, 80–97.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). "Intelligibility of normal speech. I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. 20, 255–272.

Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2002). "What's new pussycat? On talking to babies and animals," Science 296, 1435.

Calandruccio, L. and Smiljanic, R. (2012). "The Development of Sentence Recognition Materials Using a Basic Non-native English Lexicon," Journal of Speech-Language-Hearing Research 55 (5), 1342-1355.

Chung, V., Mirante, N., Otten, J., and Vatikiotis-Bateson, E. (2005). "Audiovisual processing of Lombard speech," Proceedings of the Auditory-Visual Speech Processing International Conference 2005, 55-56.

Cooke, M., and Lecumberri, M. L. G. (2012). "The intelligibility of Lombard speech for non-native listeners," J. Acoust. Soc. Am. 132 (2), 1120-1129.

Cooke, M., and Lu, Y. (2010). "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," J. Acoust. Soc. Am. 128 (4), 2059-2069.

Davies, G. M., Shepherd, J. W., and Ellis, H. D. (1979). "Similarity effects in facial recognition," American Journal of Psychology 92, 507-523.

Dreher, J. J., and O'Neill, J. (1957). "Effects of ambient noise on speaker intelligibility for words and phrases," J. Acoust. Soc. Am. 29, 1320–1323.

Fallon, M., Trehub, S. E., and Schneider, B. A. (2002). "Children's use of semantic cues in degraded listening environments," J. Acoust. Soc. Am. 121, 519-526.

Ferguson, S. H. (2004). "Talker differences in clear and conversational speech: Vowel intelligibility for normal- hearing listeners," J. Acoust. Soc. Am. 116, 2365–2373.

Ferguson, S. H., and Kewley-Port, D. (2002). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. 112, 259–271.

Garnier, M., Bailly, L., Dohen, M., Welby, P., and Loevenbruck, H. (2006). "An acoustic and articulatory study of Lombard speech: Global effects on the utterance," International Conference on Spoken Language Processing 2006, 2246–2249.

Goldinger, S. D. (1996). "Words and voices: Episodic traces in spoken word identification and recognition memory," J. Exp. Psychol. Learn. 22, 1166–1183.

Hazan, V., and Baker, R. (2011). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," J. Acoust. Soc. Am. 130 (4), 2139-2152.

Jessen, M., Koster, O., Gfroerer, S. (2005). "Influence of vocal effort on average and variability of fundamental frequency," International Journal of Speech Language and the Law 12 (2), 174–213.

Junqua, J. C. (1993). "The lombard reflex and it role on human listener and automatic speech recognizers," J. Acoust. Soc. Am. 93 (1), 510-524.

Junqua, J. C. (1994). "A duration study of speech vowels produced in noise," International Conference on Spoken Language Processing 1994, 419–422.

Junqua, J. C. (1996). "The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex," Speech Commun. 20, 13–22.

Krause, J. C., and Braida, L. D. (2002). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. 112, 2165–2172.

Krause, J. C., and Braida, L. D. (2004). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. 115, 362–378.

Lamont, A. C., Stewart-Williams, S., and Podd, J. (2005). "Face recognition and aging: effects of target age and memory load," Memory and Cognition 33, 1017-1024.

Lane, H., and Tranel, B. (1971). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. 14, 677-709.

Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F. G. (2004). "Clear speech perception in acoustic and electrical hearing," J. Acoust. Soc. Am 116, 2374–2383.

Lombard, E. (1911). "Le signe de l'elevation de la voix [The sign of the rise in the voice]," Ann. maladies oreille, larynx, nez, pharynx 37, 101-119.

Lu, Y., and Cooke, M. (2008). "Speech production modifications produced by competing talkers, babble, and stationary noise," J. Acoust. Soc. Am. 124 (5), 3261-3275.

Mattys, S. L., Brooks, J., Cooke, M. (2009). "Recognizing speech under a processing load: Dissociating energetic from informational factors," Cog. Psychol. 59, 203-43.

McCoy, S. L., Tun, P. A., Cod, L. C., Colangelo, M., Stewart, R.A., et al. (2005). "Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech," The Quarterly Journal of Experimental Psychology 58A, 22-33.

Murphy, D. R., Craik, F. I. M., Li, K. Z. H., and Schneider, B. A. (2000). "Comparing the effects of aging and background noise on short-term memory performance," Psychology and Aging 15, 49-61.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (1993). "Episodic encoding of voice attributes and recognition memory for spoken words," J. Exp. Psychol. Learn. 19, 309-328.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1985). "Speaking clearly for the hard of hearing I. Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. 28, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). "Speaking clearly for the hard of hearing II. Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. 29, 434–446.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (1989). "Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech," J. Speech Hear. Res. 32, 600–603.
Pichora-Fuller, M. K., Goy, H., and van Lieshout, P. (2010). Effect on speech intelligibility of changes in speech production influenced by instructions and communication environments. Seminars in Hearing 31, 77-94.
Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. 97, 593-608.
Pittman, A. L., and Wiley, T. L. (2001). "Recognition of speech produced in noise," J. Speech Lang. Hear. Res. 44, 487–496.
Podd, J. (1990). "The effects of memory load and delay on face recognition," Applied Cognitive Psychology 4, 47-60.
Rabbitt, P. M. A. (1968). "Channel capacity, intelligibility and immediate memory," Quarterly Journal of Experimental Psychology 20, 241-248.
Rabbitt, P. M. A. (1991). "Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ," Acta Otolaryngolica, Supplementum 476, 167–176.
Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing," Appl. Psycholinguist. 27, 465–485.
Schneider, W., Eschman, A., Zuccolotto, A. (2002). E-Prime User's Guide. Pittsburgh, Pennsylvania: Psychology Software Tools, Inc, pp. 1-160.
Schum, D. J. (1996). "Intelligibility of clear and conversational speech of young and elderly talkers," J. Am. Acad. Audiol 7, 212–218.
Smiljanic, R., and Bradlow, A. R. (2005). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. 118, 1677–1688.
Smiljanic, R., and Bradlow, A. R. (2009). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Linguist. Lang. Compass 3, 236–264.
Stollman, M. H. P., Kapteyn, T. S., & Sleeswijk, B. W. (1994). "Effect of time-scaled modification of speech on the speech recognition threshold in noise for hearing-impaired and language-impaired children," Scandinavian Audiology, 23, 39–46.
Summers, W.V., Pisoni, D.B., Bernacki, R.H., Pedlow, R.I., and Stokes, M.A. (1988). "Effects of noise on speech production: acoustic and perceptual analyses," J. Acoust. Soc. Am. 84 (3), 917–928.
Uchanski, R. M., Choi, S., Braida, L.D., Reed, C.M., Durlach, N.I. (1996). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," J. Speech Hear. Res. 39, 494–509.
Uchanski, R. M. (2005). "Clear speech," in *Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Cambridge, MA), pp. 207-235.

Uther, M., Knoll, M. A., and Burnham, D. (2007). “Do you speak E-N-G-L-I-SH? Similarities and differences in speech to foreigners and infants,” Speech Commun. 49, 1–7.
Van Engen, K., and Bradlow, A. R. (2007). “Sentence recognition in native- and foreign-language multi-talker background noise,” J. Acoust. Soc. Am. 111, 2242-2249.
Van Engen, K., Chandrasekaran, B., and Smiljanic, R. (2012). “Effects of speech clarity on recognition memory for spoken sentences,” PLoS ONE 7 (9), e43753. doi:10.1371/journal.pone.0043753